General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

NASA

Technical Memorandum 85115

PILOT CLIMATE DATA SYSTEM: A STATE-OF-THE-ART CAPABILITY IN SCIENTIFIC DATA MANAGEMENT

Paul H. Smith
Lloyd A. Treinish
Lawrence V. Novak

OCTOBER 1983

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

PILOT CLIMATE DATA SYSTEM:

A STATE-OF-THE-ART CAPABILITY IN SCIENTIFIC DATA MANAGEMENT

PAUL H. SMITH
LLOYD A. TREINISH
LAWRENCE V. NOVAK

INFORMATION MANAGEMENT BRANCH
CODE 931

OCTOBER 1983

NASA/GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND

TABLE OF CONTENTS

ILLUSTRATIONS

1.0 INTRODUCTION

NASA's Pilot Climate Data System (PCDS) offers a valuable new tool for researchers using climate-related data. The PCDS is being developed to serve as a focal point for managing and providing access to a large collection of actively used data for the earth, ocean and atmospheric sciences. The PCDS provides uniform data catalogs, inventories, and access methods for selected NASA and non-NASA data sets. Appropriate data manipulation capabilities have been developed to enable scientific users to preview the data sets using graphical and statistical methods.

Originally intended to be a part of the National Climate Program, and to provide NASA satellite data sets to outside users, the PCDS has evolved into an extensive package of capabilities to support many types of data sets from both space-borne and surface-based measurements with flexible data selection and analysis functions.

The PCDS supports various types of users. Researchers can use it to scan, analyze, manipulate, compare, display and study different climate parameters of interest. Data producers can use it for data validation, data inventory and accounting. Infrequent university-type users who may be working with a small budgets can obtain quick access to selected portions of the data. Finally, management can use it to get summary information for planning activities.

The PCDS consists of five subsystems that are described below. Briefly, they are:

1. An extensive on-line catalog that uniformly describes many data sets
2. An on-line inventory of PCDS data holdings
3. A variety of accessible data sets and a range of data set selection capabilities to select desired data according to time or geographic areas
4. A set of data manipulation utilities
5. A set of data display utilities

2.0 CATALOG

The purpose of the PCDS Catalog is to provide a central source of information about a variety of data sets in a standard format. It can be used to locate data and to determine its availability.

The Catalog describes climate-related data sets and their associated sensor measurements. At the present it includes information on more data sets than are actually contained within the PCDS. The data descriptions include the characteristics, processing status, availability and the names of people to contact for further information. This information is at a fairly high level of aggregation (e.g., all Backscatter Ultra Violet (BUV) radiance measurements from the Nimbus-4 satellite could comprise one data set), but is sufficient to enable a user to determine whether to retrieve and use data from the data sets. It describes existing data sets as well as planned data products, and currently provides descriptions for over 200 climate-related data sets.

The Catalog is completely accessible on-line through computer terminals and is occasionally printed in a hardcopy format. The on-line version consists of a summary section which can be queried using keywords, and a detailed section which can be browsed like a book. The summary section can provide a very compact output of one line per data set, or it can provide a more detailed description with a full CRT terminal screen of information per data set. The detailed section can then be browsed to get more information about any data set.

A hardcopy utility is also available for use with the Catalog and the Inventory. It allows the user to save the contents of any terminal screen for later printing or browsing on-line.

The following sources and parameters are covered by the current contents of the catalog:

SOURCES

Nimbus-4
Nimbus-5
Nimbus-6
Nimbus-7
FGGE
SEASAT
SAGE
OSTA-1
GOES (1-6)
SMS-GOES (1-2)
NOAA missions (1-7)
TIROS-N
NMC analyses
LANDSAT
AEM-2
ITOS-1
WMO surface stations
NOAA analyses

PARAMETERS

Cloud cover
Ozone
Radiation budget
Stratospheric aerosols
Wind speed
Weather variables
Ice sheet
Precipitation
Sea ice concentration
Sea surface temperature
Solar flux
Snow coverage
Nitrogen dioxide
Surface pressure
Temperature
Wave height
Sea surface elevation
Geopotential height
Albedo
Chlorophyll concentration
Humidity

3.0 INVENTORY

The main purpose of the Inventory is to provide detailed information about where to find specific data. This information is used by the Data Access Subsystem to automatically transfer data from the PCDS tape library to a disk data set when the user specifies a data type, time range, and geographic area. In this way, the mechanics of accessing a data subset are isolated from the users. The Inventory Subsystem allows a user to query a data base containing detailed information about the temporal coverage and data volume of data sets which are only available via the PCDS or closely-allied systems. The Inventory describes the data holdings of the PCDS in sufficient detail to enable the PCDS to retrieve, locate and access the requested data. An inventory graphics utility is also available to provide graphic information about data coverage and data rates.

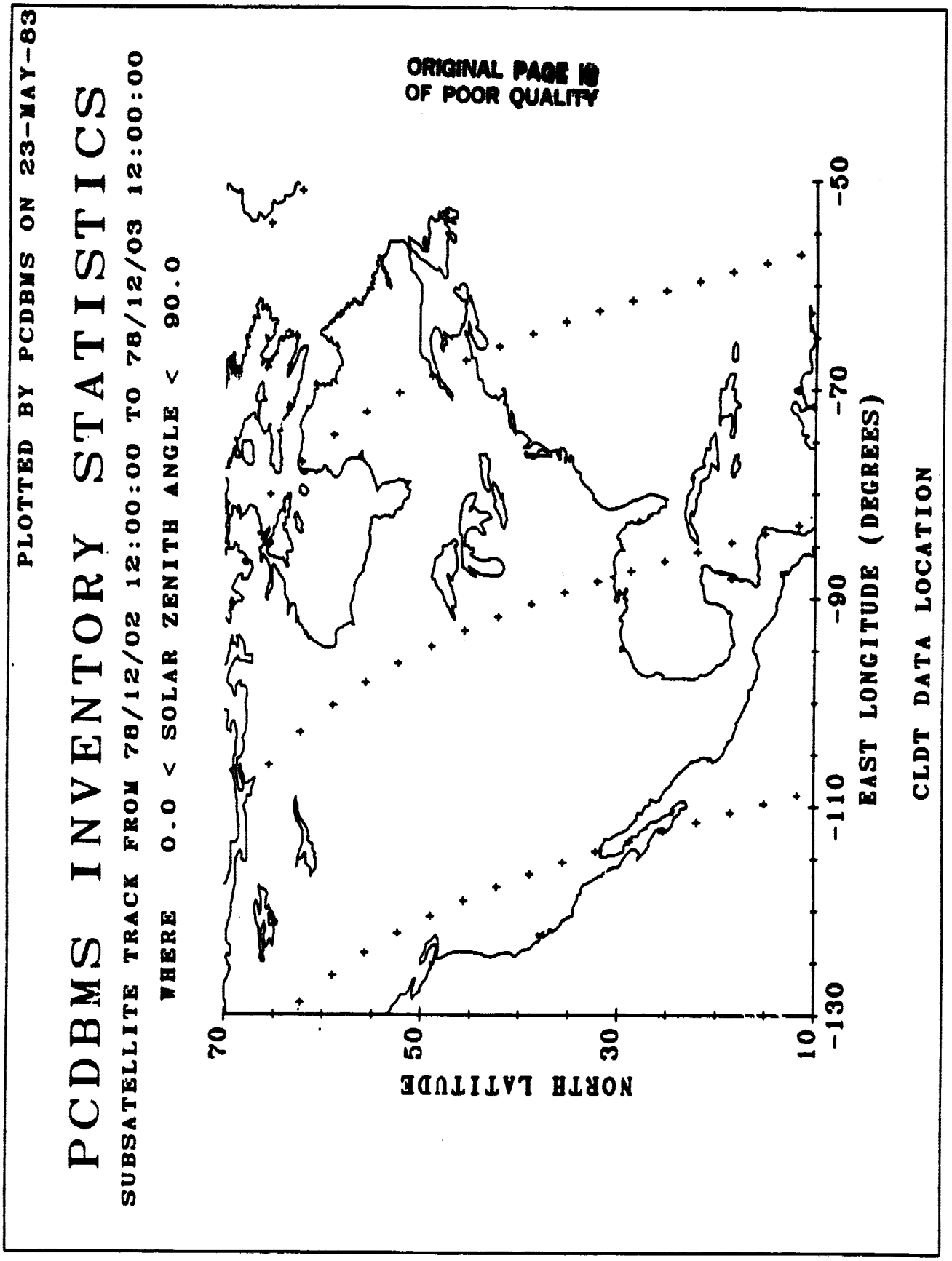
The Inventory is the lowest level in the data description hierarchy. It describes data sets at a fairly low level of data aggregation (e.g., an image, orbital strip, or physical file could comprise one data set). All data sets described in the inventory currently exist. (No planned or future products are described in the inventory.) Since the Inventory is maintained on-line in a codified format, a user can easily query it, specifying keywords to limit the information listed.

The Inventory Subsystem consists of several programs which allow a user to query the on-line inventory. Output is provided in tabular listings or a graphical format. The queries have been designed to allow a user to progress from simple queries to more detailed ones by using the information returned from the simpler queries as inputs to the more detailed queries. Another query is also available to identify additions to the inventory since a specified date.

The current capabilities of the Inventory include:

- Listing all available climate parameters
- Listing available data types and corresponding climate parameters
- Listing information about each selectable logical unit of data (e.g. maps, grids, profiles, etc.)
- Listing a summary of information about available data types, such as number of tapes, number of files, orbit range covered and the time range covered
- Listing information at the tape level about the available data, providing tape identification, number of files, orbit range, generation date, inventory date and time range
- List information at the file level about the available data, providing tape identification, file number, time range, orbit range and size.
- Listing a history of the tapes inventoried in the system
- A hardcopy utility identical to the one used in the catalog
- Graphically summarize the inventory contents by using plots of data rate as a function of time or by showing data coverage overlayed on a world map (see Figure 1)

Figure 1: Inventory Data Coverage Map



The Inventory currently supports selected data sets from the following experiments:

- Nimbus-4 Backscatter Ultraviolet (BUV)
- Nimbus-7 Solar Backscatter Ultraviolet (SBUV)
- Nimbus-7 Total Ozone Mapping Spectrometer (TOMS)
- Nimbus-7 Temperature Humidity Infrared Radiometer (THIR)
- Nimbus-7 Earth Radiation Budget (ERB)
- Nimbus-7 Stratospheric Aerosol Measurement (SAM II)
- Stratospheric Aerosol and Gas Experiment (SAGE)
- National Meteorological Center (NMC) Analyses
- World Monthly Surface Station Climatology
- First Global Atmospheric Research Program Global Experiment (FGGE)
- NOAA Heat Budget
- Nimbus-5 Electrically Scanning Microwave Radiometer (ESMR)

4.0 DATA ACCESS

The purpose of the Data Access Subsystem is to allow selection of data subsets based on time or geographic location. Data sets are kept in a library of magnetic tapes within the PCDS. These tapes are in the original format provided by the data set producer. The Data Access Subsystem allows a user to select a subset of any data set and output it in one of several formats.

The user can copy a data subset to another magnetic tape or to disk, or can simply get a hardcopy listing. The input data can be a standard tape from the PCDS library, a tape that was created earlier using this utility, or it can be a subset that was earlier placed on a disk.

One of the most useful output options is to convert the data subset into a self-describing data-independent format which can then be used by the Data Manipulation and Graphics Subsystems of PCDS. This format is called a Climate Data File (CDF). All of the data-independent software use CDF's for input and output. In addition, users of the PCDS can migrate data to their own computer facilities in the CDF format for custom analysis and return the results as a CDF for further manipulation and display.

The Data Access Subsystem employs the PCDS Inventory to automatically locate data without the user being aware of the details. The user never has to know anything about tapes, files, data formats, etc. The inventory and special software take care of all the details that are required to locate, read, and translate each of the different data types. To provide this ease of use, the Data Access Subsystem supports only those data sets that are held within the PCDS library. The data sets currently supported with Data Access are the same as those listed in the Inventory section above. (Note, however, that any data set can be put into a CDF format and used with the Data Manipulation or Graphics Subsystems. It is not necessary for a data set to be held in the PCDS library before it can be used with Data Manipulation or Graphics.)

5.0 DATA MANIPULATION

The Data Manipulation Subsystem provides a set of utilities that can be used with any data set that is in the CDF format. These utilities allow the user to customize a data set before working with it in the Graphics Subsystem, or before copying it to magnetic tape for transfer to another computer for further work. All of these utilities are designed to input a CDF and output another CDF so that they can be concatenated to perform any desired combination of functions. The utilities currently available are:

- o CDF Subsetting - to further subset data. One of the most powerful utilities provided with the PCDS is the CDF subsetting utility. It allows the customization of an output data set by selecting any fields from a CDF and editing (filtering) the data set based on any combination of fields within the CDF.

- o CDF Merge - to merge two data subsets together. One important application of this utility is to produce overlaid data sets for comparison.

- o CDF Listing - to get hardcopy listings of any data subset. The format of the listing can be tailored by selecting only the fields of interest from the CDF.

- o Ungrid a gridded data set. A utility is provided to split a map grid into its data, latitude and longitude components and place them in an ungridded format in a new CDF to permit comparisons with other ungridded data.

- o Combine CDF elements algebraically. This utility produces a new CDF with additional fields based on a user-specified combination of the fields within the original CDF. Any algebraic combination can be specified.

- o Apply statistics to CDF elements. Currently this utility supports the calculation of means and variances from any time period within a CDF.

6.0 GRAPHICS

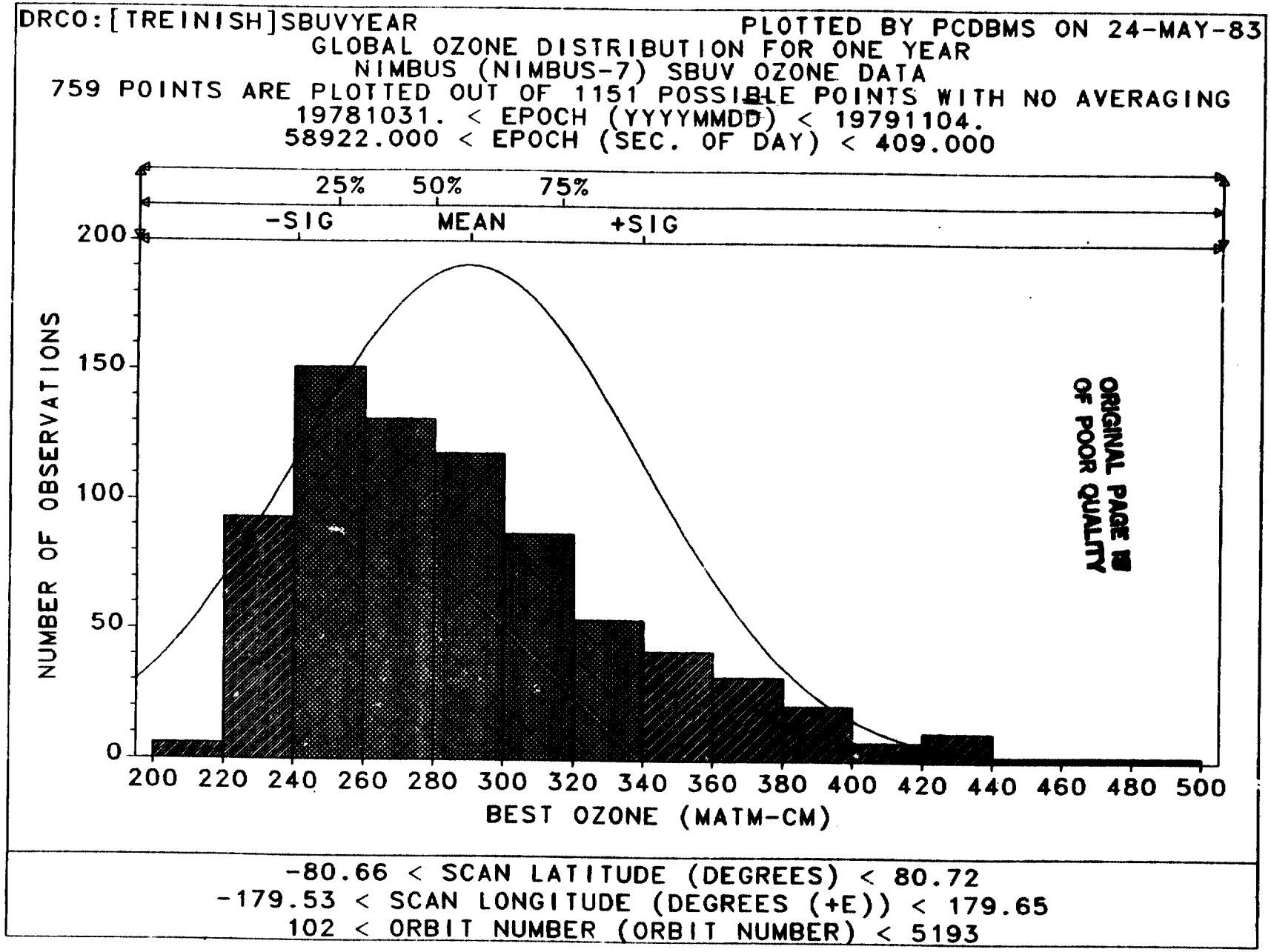
The Graphics Subsystem provides graphical display utilities for use with any data that is structured in a Climate Data File (CDF) format, whether or not the data was obtained from the PCDS. All data processed by this subsystem is stored and used in the data-independent CDF format. This subsystem can be used to preview a data set, or it can be used to get hardcopy output for further analysis, or for publication. Two- and three-dimensional graphics are supported.

The two-dimensional graphics utilities provide for standard x-y plots with one or more dependent variables on the same plot. Scatter diagrams, vector plots, and histograms are supported. Plots can be made using rectangular or polar coordinates with linear or log axes. In addition, linear, polynomial, or spline curve fits can be performed on the data set before plotting. Several types of statistics can be performed, including data smoothing and calculating the means and variance of the data. Several different type fonts are available for output, including several publication quality fonts. Figures 2 and 3 show examples of two-dimensional graphics output.

The three-dimensional graphics utilities provide contour plots, surface diagrams, and pseudo-color images. Map grids with outlines of the world coastlines in various map projections can be overlaid on these plots. These utilities also provide a large number of options to tailor the output to specific user requirements for format, quality, and content. Examples of typical three-dimensional graphics products are shown in Figures 4 and 5.

Finally, there is a graphics post processing utility which permits redirecting any output from the above procedures to various hardcopy devices. Depending on the user's requirements, options are provided for such things as publication quality output or combining several plots on one page.

Figure 2: Sample Histogram with Statistics



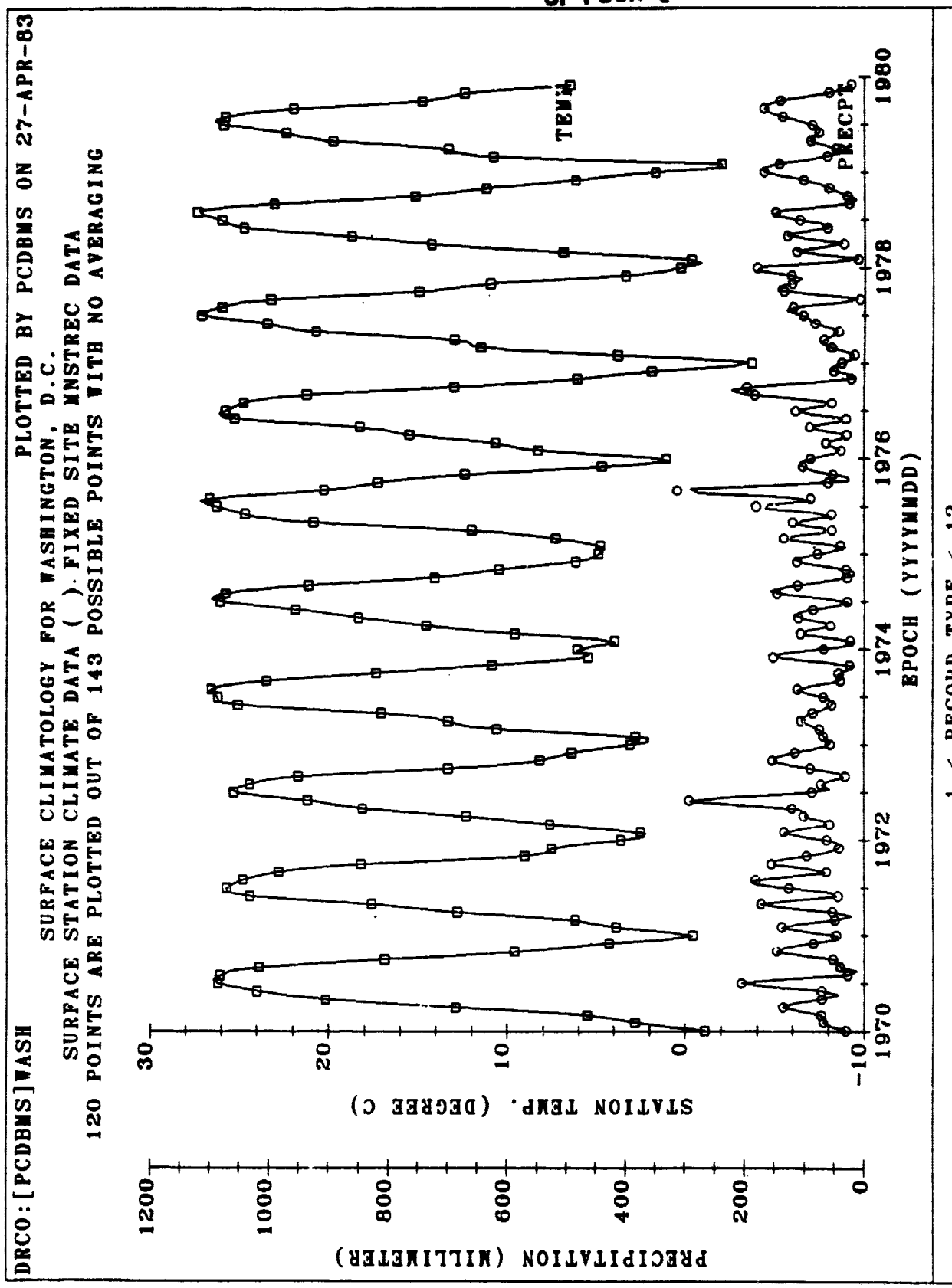


Figure 3: Scatter Diagram with Spline Curve Fit

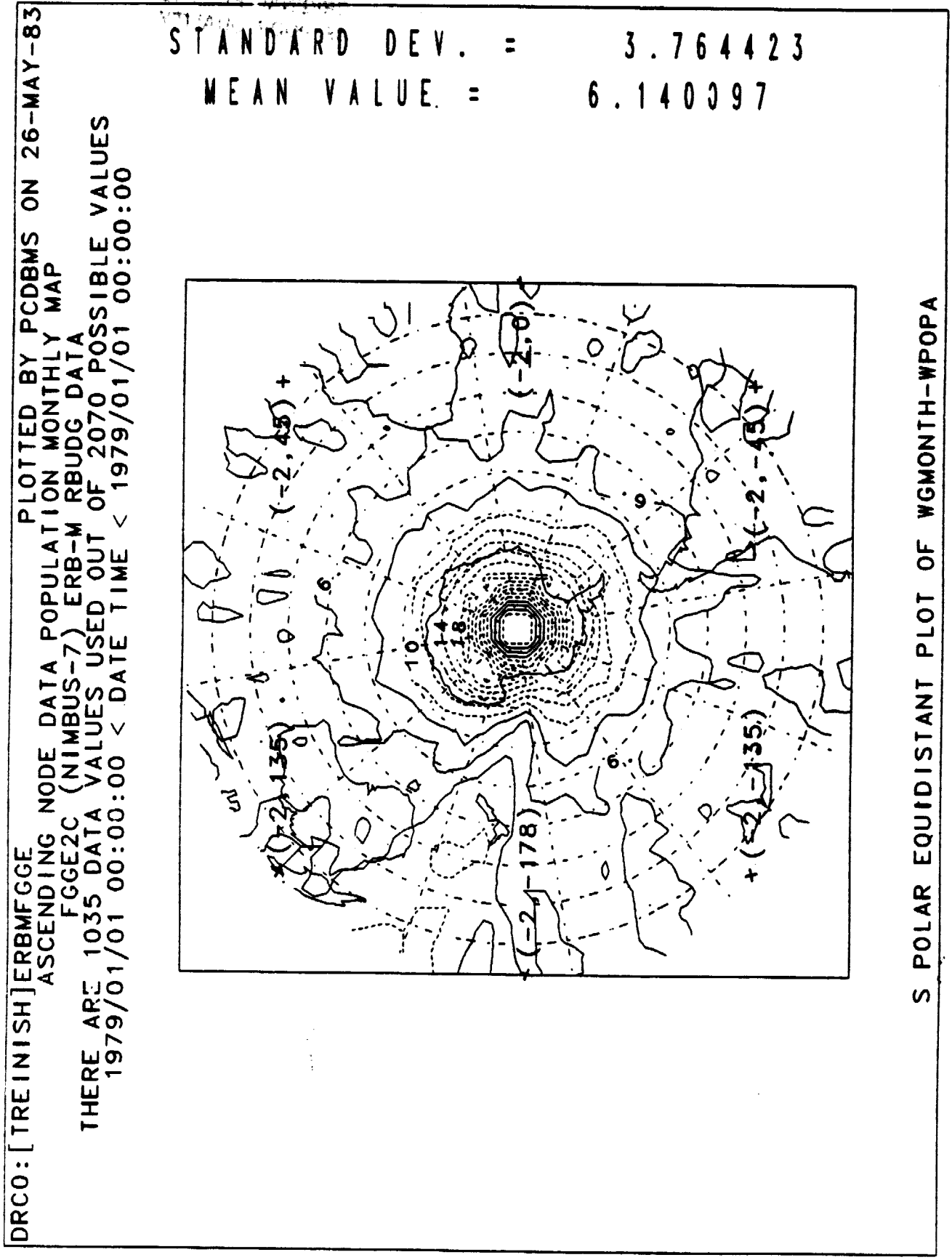
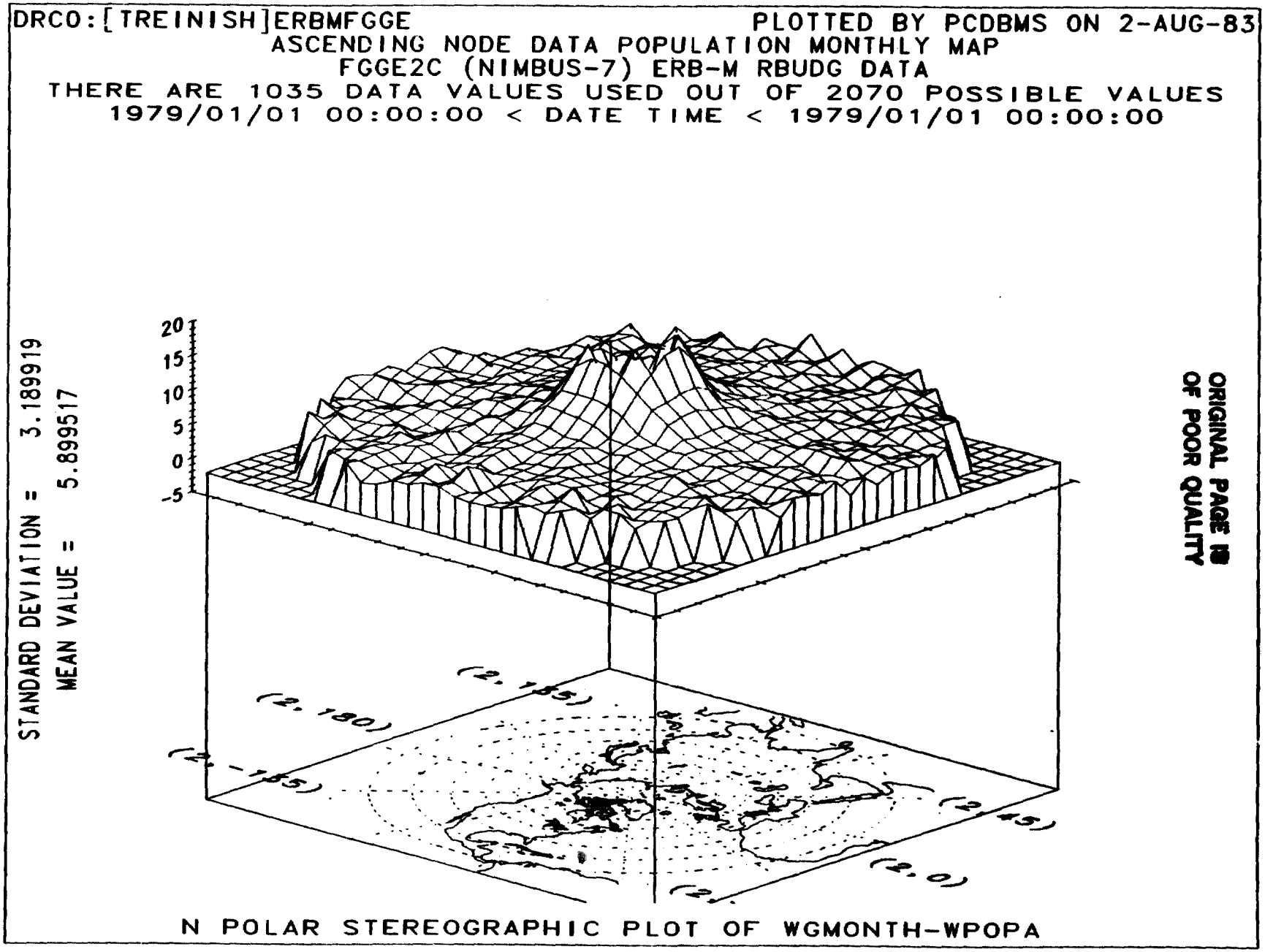


Figure 4: Contour Plot with Map Overlay

Figure 5: Surface Diagram with Map



7.0 THE PCDS IMPLEMENTATION

The PCDS is implemented on a VAX-11/780 computer system located in Greenbelt, Md. It uses the Transportable Applications Executive (TAE) as a user interface. The TAE was also developed at GSFC and provides a convenient vehicle for program development and operation. It provides a menu-driven interface for new users, and a command language interface for experienced users.

The PCDS was designed for cost effectiveness by taking existing technology and integrating it into a useful data management facility. In addition to the TAE, the PCDS uses a commercial data base management system (ORACLE), graphics package (TEMPLATE), statistics package (PROTRAN), and network communications (DECNET). By structuring the system properly, new technology can be continuously incorporated as it is developed. By taking this approach, the PCDS was able to meet its goal of demonstrating a significant application of state-of-the-art data management techniques without spending unnecessary time or money developing special purpose hardware or software.

8.0 CONCLUSION

NASA's Pilot Climate Data System provides extensive data management and analysis facilities for use by climate researchers. These facilities are packaged into a system that is extremely easy to use and very flexible. The use of the PCDS requires no knowledge of data formats or programming languages, nor does it require any special programming to support a given application.

The PCDS has provided direct support to climate researchers at Goddard Space Flight Center. The system is constantly being improved based on the requirements of the supported users. At the same time, the support of new data sets that users request and general purpose capabilities based on user requirements are always being added. In addition, research and development activities are conducted in data management techniques and the results are applied to improve the PCDS.

The PCDS has evolved from the demonstration of state-of-the-art data management techniques that were applied to support scientific research using large data sets from many different sources. In addition to its practical applications, it is expected to serve as a model for future, larger, data management systems which will be part of still larger facilities which will serve many different scientific disciplines.